CONTENTS

Report Number	Title	Author
1.	Storage and Retrieval of Contents of Technical Literature, May 1956	Don D. Andrews et al.
2.	Advances in Mechanization of Patent Searching, April 1956	B. E. Lanham et al.
3.	Problems in Mechanizing the Search in Examining Patent Applications, October 1956	Simon M. Newman
4.	Storage and Retrieval of Contents of Technical Literature, June 1957	Simon M. Newman
5.	Organization of Chemical Disclosures for Mechanized Retrieval, June 1957	B. E. Lanham et al.
6.	(ILAS), June 1957	Don D. Andrews
7.	A Punched Card System for Searching Steroid Compounds, July 1957	J. Frome et al.
8.	Recent Advances in Patent Office Searching: Steroid Compounds	J. Frome et al.
	and ILAS, 1957	Don D. Andrews
9.	Linguistic Problems in Mechanization of Patent Searching, 1957	Simon M. Newman
10.	A First Approach to Patent Searching Procedures on SEAC, January 1958	H. Pfeffer et al.
11.	A Manual for Coding Steroids, November 1958	J. Frome et al.
12.	Storage and Retrieval of Contents of Technical Literature, November 1958	Simon M. Newman
13.	A System of Retrieval Compounds, Compositions, Processes and Polymers, November 1958	J. Frome et al.
14.	Variable Scope Patent Searching by an Inverted File Technique, Nov. 1958	J. Leibowitz et al.

CONTENTS

Report Number	Title	Author
15.	A Notation System for Transliterating Technical and Scientific Texts for Use in Data Processing Systems, May 1959	Simon M. Newman et al
16.	Analysis of Prepositionals for Interrelated Concepts, July 1959	Simon M. Newman
17.	Semi-automatic Indexing and Encoding, December 1959	J. Frome
18.	Mechanized Searching of Phosphorus Compounds, January 1961	J. Frome
19.	Revised Steroid Search Systems Coding Manual, January 1961	J. Frome
20.	Parameters for an Information Retrieval System for Chemical Processes, April 1961	R. W. Swanson et al.
21.	A New Sequential Enumeration and Line Formula Notation System for Organic Compounds, November 1961	H. W. Hayward
22.	Manual for a Punched Card Retrieval System for Organic Phosphorus Compounds, November 1961	J. Frome et al.

Patent Office
Research and Development
Reports

STORAGE AND RETRIEVAL OF CONTENTS OF TECHNICAL LITERATURE

NONCHEMICAL INFORMATION

Preliminary Report
May 15, 1956

FOR OFFICIAL DISTRIBUTION

Prepared by

Don D. Andrews
Director

Simon M. Newman
Patent Research Specialist
Staff Member

Office of Research and Development
U. S. Patent Office

Robert C. Watson
Commissioner of Patents

Sinclair Weeks
Secretary of Commerce

OFFICE OF RESEARCH AND DEVELOPMENT REPORTS

Errata

Storage and Retrieval of Contents of Technical Literature

Non-Chemical Information

Preliminary Report

p 3, col 1, l. 37	for "clips", read -chips-
l. 44	for "principal", read -principle-
p 5, col 1, l. 35	for "empackaged, read -enpackaged-
p 6, schedule 2,	meaning 4, under Examples of Use, last example, insert -thru- after "book"
p 7, schedule 6,	under Explanation, ll. 1 & 2, ll. 3 & 4 and ll. 5 & 6 should be interchanged.
schedule 7,	under Concepts and Interfixes ll. 2 & 5 should be interchanged.
p 9, col 1, l. 22,	for "except" read -concept-
p 11, schedule 15,	the title Descriptor and Modulant should be supplied for col. 2.
, col 1, l. 17,	after "item" insert -below as-
schedule 17,	under the title Concept and Interfix, the numeral -4- should be inserted in ll. 2, 4 and 6 immediately above the "4" in ll. 7, 9, 10 & 12.

Problems in Mechanizing the Search in Examining Patent Applications

p 6, col 1, l. 23	for "sytem" read -system-
p 15, col 1 between l. 7 & 8,	insert, -which might have any of the details for-
p 21	the title, "Exhibit 9" at the bottom of the page should appear directly below the illustration at the top of the page.

STORAGE AND RETRIEVAL OF CONTENTS OF TECHNICAL LITERATURE

NONCHEMICAL INFORMATION

INTRODUCTION

This is the first of a series of papers reporting the results of research directed to the storing and retrieval of scientific information disclosed in the contents of non-chemical patents and other technical documents. This paper constitutes a preliminary report of the work done to date, and of our thinking at this time. Changes and developments will be reported in future papers.

This research is a second step by the Office of Research and Development, (U. S. Patent Office) in its study of methods of storing and retrieving information in scientific disclosures. A chemical task-force has already reported (1) on the development of a proposed system for handling chemical disclosures. All of the research in this field done to date by the chemical task-force has been carefully studied and used by us, and credit for many of the ideas here propounded is freely given them.

In any system which may result from this research, there are at least two features which will be desirable. First the system, in so far as possible, should be compatible with the system under development by the Chemical task-force which will be used to search the Chemical Patents and Literature. And secondly, it should be able to encode for retrieval, any disclosed feature of the document being encoded. Policy will later dictate what features, if any, should be omitted in the encoding process.

However, in determining what portions should be omitted when encoding a document, it must be kept in mind that many processes disclose a by-product which may not seem important or even relevant at the time of coding. A very specific machining operation upon a metal blank to form a particular machine part may incidentally leave a pile of intertangled metal clips, which are disclosed as mere scrap. However, this disclosure may be the very reference which will be later wanted in a search for the manufacture of steel wool, and failure to encode the scrap as one product of the machining operation, if the other product is encoded, can be seen to be an error in principal.

Since most disclosures of scientific information are already linguistically expressed, and it seems clear that others in the form of drawings, tables, photographs, models, working machinery, etc. are translatable into language form, it was felt that a purely linguistic approach to this problem warranted investigation. As this particular study progressed we slowly came to the conclusion that we could not use either the word position in a sentence, or the grammatical construction of the sentence in the solution of this problem.

"LANGUAGE" TO BE USED

Professor Stuart C. Dodd, in his paper "Model English"(2) has suggested the development of a "Ruly" English (as opposed to the common "unruly" language which we now have) in which every word would have one, and only one conceptual meaning, and in which each and every concept would have only a single word to describe it. Our study contemplates, at least in part, the creation and use of such a Ruly Language—however a specialized one, wherein the "words" will be designed and adapted for information storage and retrieval, and not necessarily styled for conversation or writing.

It is unfortunate that in the English language there has been no uniform or logical rule for the naming of devices or things. A few things are named for their shape, for example, a block or a ring. Others are named for the material from which they are made, for example, a glass. The great bulk of things which we refer to are given functional names because of the process they perform, for example, a press or a hammer; or for the use to which they are put, for example, a receptacle or a cover. Others are named for the location from which they first came, for example, china. Some few have arbitrary names, for example, a pencil or an ax. Thus, except for those named for their shape (which constitute the only words truly descriptive of the static structure of a thing) and those named arbitrarily, we see that the names, themselves, are in reality, either broad relationships with other things which are not recited or broad statements of processes of use of the thing.

In the absence of a Ruly English, we may have to use several words joined together by hyphens(-) to simulate a Ruly English word in examples of our coding system.

A SIMPLIFIED EXAMPLE OF ENCODING

An analysis of the data needed to be encoded for retrieval indicates that such data is reducible to the recitation of either (1) named things alone, or with descriptive explanations thereof, or (2) two or more of these "named things" with a stated relationship of each to one or more of the others.

Let us consider a very simple disclosure, viz: A table standing on the floor, and an ash tray and book spaced alongside each other on the table. The table constitutes a support, it has plural legs, its top is flat, and it supports a book and an ash tray. The book has a leather cover, it is colored blue; the ash tray is made of metal, is used as a container and lies adjacent to the book.

Each of the terms enumerated in this disclosure which recites a characteristic of one of the disclosed things may be translated into a Ruly word. Upon analysis, it will be seen that each such word describes the thing from one aspect.

We will accordingly group the words which characterize each thing. Additionally, where a specific interrelation between two or more things is stated, we will place into *each* group, one of a set of cognate words, which words together will express this interrelation.

These cognate words will be "mirror-images" of the explicit interrelationship expressed between the things described by the words of the two groups and will act to couple the groups by this relationship. E.g., if A supports B we will put "supporting" with A and "supported-by" with B. If P is greater than Q, we will put "greater-than" with P and "lesser-than" with Q. And if M is equal to N, we will put "equal-to" with both M and N.

We will also place an arbitrary but identical number with each cognate word to indicate which words constitute the interrelated couple between the groups. A second set of cognate words describing a second and different relationship, i.e., a different couple, will therefor carry different, arbitrary, identical numbers.

For convenience of expression, each individual word will be called a "descriptor" and the group of descriptors relating to one named thing will be called an "item." Each of the items will be numbered consecutively, only for ease in referring to them, since it will be later shown that their order is not material. The process of showing interrelational concepts of two things by placing descriptors of their interrelation in each of the items so related will be called "distribution." The arbitrary number indicating the parts of a distributed concept to be coupled will be called an "interfix." Since the numerical value of the interfix merely shows identity of the distributed concepts, the retrieval of such distributed concepts can only be made on requests for identity of the interfix. This type of retrieval will be called "blind retrieval."

Item #	*Descriptors*	*Interfix*
100	Table	
	Support	
	Multi-legged	
	Flat-top	
	Supported-by	1
	Supporting	2
101	Floor	
	Supporting	1
102	Book	
	Leather-binding	
	Blue-colored	
	Supported-by	2
	Adjacent-to	3
103	Ash-tray	
	Container	
	Metallic	
	Supported-by	2
	Adjacent-to	3

SCHEDULE 1

The disclosure recited above might be illustrated as shown in Schedule 1.

In the preceding example, we see four items, 100 listing the descriptors of the table, 101 the floor, 102 the book and 103 the ash tray. By the use of the separate descriptors in each item, we have described each thing as a whole by each of its various disclosed descriptive characteristics.

By the use of the interfix #1 we have linked together the descriptor "supported-by" relating to the table (100) and the descriptor "supporting" of the floor (101), designating the supporting of the table on the floor. In the same manner the interfix #2 designates the book and ash tray supported by the table and interfix #3 designates that the book and ash tray are each "adjacent-to" the other.

It can thus be seen that the things itemized may be retrieved either generically or in detail, and in either combination or subcombination, with or without their interrelations. Also, and most importantly from the Patent Office point of view, the interrelationships there recited may be searched for and retrieved, independently of the details of the specific things recited. For example, not only can we retrieve a metallic ash tray supported upon a multi-legged table, but we can retrieve a container upon a support or a blue thing and a metal thing adjacent to each other, whether or not they have a common support.

This system may also be utilized in process or method disclosures, an example of which will be given later (Schedule 9).

Some refinements which have been considered will now be discussed. None of these can be said to have been completely debugged.

USE OF WORD ROOTS WITH MODULANTS

We have previously pointed out that the names of things often state various relationships. Upon analysis, both names and other words used as descriptors usually infer either a broad relationship with some other unidentified thing, or an indefinite or undefined relationship with the specific thing being itemized. E.g., in Schedule 1 the word "container" in item 103 indicates that the thing (ash-tray) itemized in 103 is a "container" or "holder" for something not itemized, while the "leather-binding" of item 102 indicates that the thing (book) itemized in 102 has previously been thru some indefinite and undefined process called "binding."

Let us postulate a Ruly word ENPACKAGE, defined in unruly English as the process of surround-

ing some other thing with an enclosure, wrapper or container to form a package. This concept has been clearly, and we believe unambiguously, stated elsewhere(3), and includes within its scope, the process of enclosing a candy bar in a tin-foil wrapper, filling a preformed paste-board box with cereal and closing the box, or enclosing loose tea in permeable material to make a tea bag.

In addition to (1) "the process" enpackage, there come to mind additional corollary concepts such as—

(2) The "enpackager" or package-maker, i.e., the performer of enpackage, known in the Patent Office as "the apparatus."

(3) The thing-made-by-enpackage, i.e., a package, known in the Patent Office as "the final product" or more simply, "the product."

(4) The filling-(or contents-)-which-was-"enpackaged," i.e., the candy bar, the cereal or the tea, known in the Patent Office as the "starting material," or sometimes, in other types of processes as "the stock material."

(5) The partially-completed-product-at-some-point-during-enpackage, i.e., the candy bar partially wrapped, or the full but still open cereal box, known in the Patent Office as "the intermediate product," or sometimes, in other types of processes as "the blank."

(The product (3), the starting material (4) and the intermediate product (5) are all part of a genus known in the Patent Office as "the work," in other words, the-thing-worked-on-in-enpackage(or by-the-"enpackager")).

(6) The-condition-of-being-"enpackaged" as distinguished from being "unempackaged" or loose.

(7) Another-thing-made-from-a-product, e.g., a cup of tea made from a tea bag. (This thing, of course, is a product of another process—the-process-of-making-tea.)

(8) A-larger-thing-of-other-characteristics-of-which-the-product-is-a-part, e.g., a merchandising display of a product. This would be known in the Patent Office as a combination including the product as a subcombination. Note that concept (7) is a genus of which this concept (8) is a species.

We can now take the root of our ruly word enpackage, viz ENPACK-, and by means of a series of inflecting codes, which we call "modulants," build the concepts of—

(1) Process
(2) Apparatus
Work
(3) Product
(4) Starting-material
(5) Intermediate-product
(6) Condition
(7) Made-from
(8) Combination-including

where the species of a genus is shown by indenting the species under its genus.

As pointed out later, we may want to include as a descriptor, the common name of the thing being itemized. For want of a better means of identifying these non-ruly names of things, we shall use in the examples which follow a modulant called "named-thing."

Other modulants will easily come to mind. It seems clear that a series of modulant codes can be devised with which descriptor word-roots can be modulated to show the particular concept present in a disclosure to be encoded. Such a code has not yet been worked out in detail.

These modulants are inflectors of roots to allow them to serve as descriptors, and modulation is desirable because the root form may be retrieved without the modulant when making a generic search.

RULY ROOT CODING

The coding of the unmodulated roots presents a linguistic problem of no mean complexity. Creation of generic hierarchies of these roots which will be meaningful is necessary. The term "steam" as a process, is specific to "evaporate," as a product it is specific to both "water" and "fluid" while as an apparatus, it is generic to a "steam generator" (i.e., a boiler). In the beginning, we contemplate the necessity of utilizing a plurality of hierarchial codes for each root, and as our coding develops, we may be able to combine some of them and reduce their number.

THE INTERRELATIONAL CONCEPT

As pointed out before, relations between items are encoded by means of coupling descriptors with interfixes. In other words, in addition to modulating the root word to form a descriptor the interrelation must be shown by some additional inflection of that word. These inflectors will be called "Interrelational Concepts." They are often prepositional in form.

Early in this effort, it was discovered that the meaning of any preposition was very elusive and ambiguous. 25 *Basic English* words which were recognized as prepositions (about, across, after, against, among, as, at, before, between, by, down, for, from, in, of, near, off, on, over, thru, till, to, under, up and with) were accordingly studied with an attempt to discover what basic concepts they portrayed.

Each of these prepositions was then further analyzed for as many of its unambiguous meanings as could be found. Each meaning was given a number which we call a "meaning number" and in lieu of definition, an equivalent expression and examples of its use were listed. (In this work, Mr. C. G. Smith of this Office, gave considerable time and advice). An example of the breakdown of the preposition "thru" follows:

THRU

Meaning #	*Equivalent Expression*	*Examples of Use*
1	Between the boundaries of	A hole extending thru a board
2	Between and across the boundaries of	The route of work thru a machine, An arrow extending thru an apple
3	Everywhere in	An odor pervading thru a Room; Poison thru the dead body
4	Progressing from beginning to end of	To pass thru a doorway; A bill passing thru the legislature; To read a book from cover to cover
5	On a route between portions of	The flight of an arrow thru the air; A ray of sunlight shining thru the trees
6	Here and there in	Pollen floating thru the air; Raisins scattered thru the bread dough
7	By the way of	To cure thru an operation; To change thru legislation
8	Because of	The water froze thru loss of heat; To err thru ignorance
9	By use of	A solution thru calculus; To speak thru an interpreter
10	Be finished with	At 6:00 P.M. he was thru with work
11	Something simultaneous with and despite of	To hear thru the din; To see thru the fog, To see thru his deceit
12	During and to the end of	Busy all thru the year; To go thru life without ever knowing
13	By the mere existence of	To be related thru marriage

SCHEDULE 2

Pending the coining of a Ruly word we distinguish the different concepts of the same unruly word by writing after the word the meaning number which we have assigned, thus: He was thru(10) with his homework. The sketch was first done in(8) outline.

These breakdowns of the 25 prepositions were then scanned for redundancy, and a series of Ruly Words were constructed for some of the concepts. Equivalent prepositional phrases were also noted as we ran across them, their separate meanings were numbered and equivalent expressions and examples were listed. As an example of such a Ruly Word, we have the concept Howby, which has the meaning, *mode of proximate cause*, and has equivalent meanings "resulting from the mode" and "by the use of the mode." The terms collected from each of the pertinent breakdowns are shown in Schedule 3.

With this analysis, we believe we have tied down the meaning of "howby" to a single concept.

Next these Ruly concepts were organized into generic hierarchies, the most generic being placed farthest to the left, the subgeneric indented thereunder and the most specific farthest to the right. For example, the several categories of Cause follow:

CAUSE, the proximate cause of a Result

Ruly Word	*Explanation*
CAUSBY	Proximate cause
HOWBY	Mode of causing
MEANSBY	Means for causing

SCHEDULE 4

As stated previously interrelational concepts are complementally inflected into mirror-images of the concept, when distributed. E.g., in the case of cause, the thing causing an effect results in a complementary result. E.g., "a cut from a knife" could be itemized in unruly English as:

Item #	*Descriptors*	*Interfix*
201	Cut	
	Resulting-from	7
202	Knife	
	Caused-by	7

SCHEDULE 5

HOWBY

Term and Meaning #	*Examples of Use*
Prepositions	
As(3)	—To limp as the result of a fall
By(14 part,* 18 & 21)	—To take by force; To teach by example
From(10 part*)	—To gain a polish from wear
In(8)	—To sketch in outline; To argue in a circle
Of(25 part*)	—To go of one's own will
Thru(7, 8 & 9 part*)	—To cure thru an operation; To gain thru legislation; a solution thru calculus; To freeze thru loss of heat
To(43)	—To succumb to a force
With(7 part*)	—To kill with kindness
Phrases	
Because-of(2)	—To acquire polish because of wear
Result-from(2)	—A cure resulting from an operation
By-use-of(1)	—A solution by use of calculus

*By building single concepts, we discovered that many of the prepositional meanings we thought were unambiguous still covered more than one meaning, and we had to split the meaning numbers already assigned. Hence in by(14) we used only some of the examples, and left the others for another concept. We also noted occasions where an unambiguous meaning had been given several meaning numbers and we therefore combined by(18) and by(21) with part of by(14) in making Howby.

SCHEDULE 3

CODING OF CONCEPTS

In order to handle these concepts in "item" form, and at the same time to utilize the generic hierarchy form, a specific system of coding them was adopted, which we call "compliance coding."

These compliance codes are binary in form, that is, the presence or absence of qualities is noted by 1's and 0's in separate columns. These codes are organized and arranged so that the most specific concepts have the fewest 1's in their code. Subgeneric concepts add additional 1's in other bit columns, and the most generic code includes a 1 in every column in which a subgenus or species exists.

The coding of the Cause-Result concept, with addition of Ruly words to enable one to easily refer to the codes follows:

Ruly Word	*Code*				*Explanation*
CAUSBY	0	1	1	1	Result of
CAUSFROM	1	1	1	0	Caused by
HOWBY	0	0	1	1	Result of process
HOWFROM	1	1	0	0	Process causing
MEANSBY	0	0	0	1	Mechanical result of
MEANSFROM	1	0	0	0	Mechanism causing

SCHEDULE 6

Causby, it will be noted, has a 1 in each of the last three columns, and a 0 in the first column. Any other word which has a 0 in the first column and a 1 in one or more of the other three columns is a species under the genus Causby. Hence Howby and Meansby are both species of Causby, but by the same rule, Meansby is also a species of the subgenus Howby. In a like manner we have the genus Causfrom, the subgenus Howfrom and the species Meansfrom.

Now we can take the example of Schedule 5 and using our ruly concepts, we have:

Item #	*Descriptors and Modulants*	*Concepts and Interfixes*
201	Cut—(product) Cut—(process)	Meansby 7
202	Knife—(named-thing) Cut—(apparatus) Cut—(process)	Meansfrom 7

SCHEDULE 7

An example of a more complex compliance code is that shown in Schedule 8.

In this hierarchy, it will be noted that neither Syncwith nor Timnear are grouped with another word. In these cases, the complementary concept terms are identical in both original and "mirror image" form. We also note that Timnear is generic to both Timafor and Timaft, though neither of the latter are subgeneric to each other.

Only a few hierarchies of interrelational concepts have been formed to date, and not all of them have been coded. A number of concepts have also been collected, for which no hierarchies have been formed. The great bulk of the concepts encompassed by the 25 words so far analyzed have not been collected. We are experimenting with a more direct approach in forming these concepts.

Ruly Word	*Code*								*Time Diagram*	*Explanation*
SYNCSTART	1	1	0	0	0	0	0	0	x	Unequal, simultaneous beginning
SYNCBEGIN	0	1	0	0	1	0	1	0	xxxxxxx	
SYNCSTOP	1	0	1	0	0	0	0	0	x	Unequal, simultaneous end
SYNCEND	0	1	1	0	1	1	0	0	xxxxxxx	
SYNCWITH	0	1	0	0	0	0	0	1	xxxxxx xxxxxx	Simultaneous
DURING	1	0	0	0	0	0	0	0	x	once / repetitive: shorter during longer
WHILE	0	1	1	1	1	1	1	0	xxxxxxxxxx	
RECURPER	1	0	0	1	0	0	0	0	x x x	
AFORLAP	0	0	0	0	0	1	0	0	xxxxxxxx	Overlapping periods
AFTLAP	0	0	0	0	0	0	1	0	xxxxxxxx	
TIMAFOR	0	0	0	1	0	0	0	1	xxxx	Before and after
TIMAFT	0	0	1	0	0	0	0	1	xxxxx	
TIMNEAR	0	0	1	1	0	0	0	1	(not illustratable)	Sequential, no sequence expressed

SCHEDULE 8

RELATION OF MODULANTS TO INTERRELATIONAL CONCEPTS

Since many, if not all, the modulants are used to show relationships, it may well be that the modulant codes, when created, will be closely related to the interrelational concept codes. This relationship has yet to be analysed.

MORE COMPLEX ENCODING—USING METHOD AS EXAMPLE

Now let us return to the itemization of a simple procedural method, e.g.: Filling a glass measure from a china pitcher and emptying the measure into a larger metal container. In the schedule which follows, the substitutes for modulants are enclosed in parentheses and the prepositional concepts, used instead of their codes, are in italic.

Item #	*Unruly-Root and Modulant*	*Ruly Concept and Interfix*		
209	Pitcher —(named-thing) Contain —(apparatus) Lip —(combination-including) China —(made-from) Dispense—(apparatus) Dispense—(method)	*fromout-1*	*syncwith-2*	*timafor-6*
210	Measure—(apparatus) Contain —(apparatus) Glass —(made-from) Size Dispense—(method) Dispense—(method)	*lesser-5* *into-1* *fromout-3*	*syncwith-2* *syncwith-4*	*timafor-6* *timaft-6*
211	Contain —(apparatus) Metal —(made-from) Size Dispense—(method)	*greater-5* *into-3*	*syncwith-4*	*timaft-6*

SCHEDULE 9

In this schedule, we find item 209 directed to a thing called a pitcher variously described as a container, a thing having a lip, a thing made of china, a thing called a dispenser, and interrelated in a dispensing process step between the pitcher and some other thing. Interfix 1 shows that this step is "out of" the pitcher and that it is "into" the measure of item 210. Interfix 2 shows that this dispensing-receiving step is simultaneous. Interfix 6 shows that this dispensing-receiving step occurs before a second dispensing-receiving process step between the measure 210 and the container 211. Interfix 4 shows that this second step is also simultaneous.

We note that the measure of item 210 is made of glass and by interfix 5, that it is smaller in size than the metal container of item 211.

SAME CONCEPT DIFFERENTLY EXPRESSED

Coding will obviously not all be done by one person, nor will the question for retrieval normally be framed by the person who did the encoding. An example of efficacy of the "interrelational except" in this situation may be interesting. Consider the simple disclosure: "The water is emptied *from* the pitcher." And suppose a search question is framed in the form: Find "The pitcher is emptied *of* its water."

Coding disclosure and question we have:

Item #	*Unruly Root & Modulant*	*Ruly Concept and Interfix*
	Disclosure	
27	water-(work) empty-(process)	fromwhence-6
28	pitcher-(named-thing) empty-(process)	whencefrom-6
	Question	
1	pitcher-(named-thing) empty-(process)	whencefrom-1
2	water-(work) empty-(process)	fromwhence-1

SCHEDULE 10

and we note that the two sets of items are identical, although the order is changed. Since the order of items is not material, we see that the question will retrieve the disclosure.

These two sets are identical because the two concepts are identical. This is preordained in view of our analysis of from(3), listed as: whence, e.g.: "He took a penny from his pocket" and *of(14)* as: out from, e.g.: "It was a wine *of* France." These were both selected as elements of the ruly word whencefrom, along with *on(18)*: out of, e.g.: "His check was drawn *on* the bank," and *off(1)*: remove from, e.g., "He cut the end *off* the stick."

SERIAL NUMBERING

The handling of complex structures, whether static or dynamic, presents further problems. A table has plural legs, each of which may need separate identification. A transmission similarly has plural gears. To take care of this situation, we propose a complex notation which we call Serial-Numbers. Like our interfix, this will involve a blind retrieval process. These numbers will be assigned so that any larger combination will have the same significant figures as each of the subcombinations which belong to it. Referring to our first itemized disclosure, (SCHEDULE 1) we might assign serial numbers as follows:

Entire-Disclosure	1	0	0	0
Table-and-Contents	1	1	0	0
Table-top-and-contents	1	1	1	0
Table, first-leg	1	1	2	0
Table, second-leg	1	1	3	0
Table, third-leg, etc.	1	1	4	0
Table-top	1	1	1	1
Book	1	1	1	2
Ash-tray	1	1	1	3

SCHEDULE 11

The use of this serial-number notation appears necessary, but the utilization of it from the standpoint of retrieval has not been studied in detail.

MODIFYING CONCEPTS IN GENERAL

The modifying concepts of English, i.e., the adjectives, adverbs and prepositional phrases can apparently all be handled by our system. Adjectives fall in several classes, each of which requires a different technique in coding.

First, there are the purely descriptive modifiers or qualifiers. These are recognizable because the sentence in which they are used may be modified by making the word a predicate adjective following the verb to be. E.g., "A cold press" is equivalent to "The press is cold." These words will be entered as modulated descriptors.

There are next those modifiers which show the role of the modified noun, or something concerning it. These are usually concepts involving another

thing and require a separate item with a distributed concept. "A power path" expresses the concept "a path for power," similarly "the Potomac Bridge," "a bridge over the Potomac," "a telephone call," "a call on (or by) a telephone," etc. As an example of the role situation, we may itemize the expression, "a cracker box" by:

Item #	*Unruly Root and Modulant*	*Concept and Interfix*
428	Box—(named-thing)	*Containing-4*
429	Cracker—(named-thing)	*Contained-in-4*

SCHEDULE 12

This method of handling qualifiers requires unambiguous contexts for encoding. "A German Book" must be encoded as either "A book in German," "A book from Germany," or "A book about Germany," according to the context.

The combination-subcombination relationship, which is a modifier of this form, presents some problems if the Serial Number notation referred to above is adhered to. As pointed out in Schedule 10, other means has been proposed for handling the expression "table-top" when referring to the specific top of a specific table. It has not yet been determined in which manner such concepts will be encoded.

QUANTIFIERS

The quantifiers, i.e., adverbial words modifying adjective words, are indications of relative position on a scale. As such they are interrelational concepts, and can be handled in that way. See, e.g., Schedule 9 where the metal container is larger than the glass measure. But as has been pointed out before, some relations are expressed generally without being interrelated to another thing. We have spoken of an ash tray as a container, without interrelating it with the smoker's waste material, and we have called the book blue, without interrelating it to a standard color chart. Hence such ambiguous statements as a "large" ash tray or a "light"-blue book can not be coded as interrelational concepts. Whether they can be coded as modulants has yet to be investigated.

INVARIABLE CODES

Certain aspects of disclosure are normally used in conjunction with another concept. These are the aspects of temperature, weight, elapsed time, volume etc. It appears clear that a code for encoding such measurable items could use numerical values preceded or followed by a fixed code notation which would mean, for example, time (in minutes), volume (in cubic meters), temperatures (in degrees Kelvin), etc. Such coding will be called "invariable coding" since no modulation or no genus—species relationships occur in these codes.

INDEX NUMBERS

Since many things will undoubtedly be searched for by common name, we will compile an alphabetical collection of common terms with index numbers for each term. Where a term has a different meaning in different arts, two index numbers will be given, for example:

BRAKE

Motion Snubber —1,000,000,071

Sheet Metal Bender — 350,896,253

SCHEDULE 13

Such common names can then be listed as a descriptor in an item, and will thus allow retrieval of such things by their common name.

SPECIAL RELATION CODING

Certain special aspects, which can be handled by some or all of the details already referred to, can also be handled in other ways, which have certain advantages and solve other problems. They have been exploited for the solution of these latter problems, and it is possible that future research may generalize on these techniques for other and different problems.

Many of the interrelationships searched in the Patent Office have a dominant-recessive character, others are equi-relative. E.g., a tractor (dominant) pulls a trailer (recessive), a table (dominant) supports a book against the pull of gravity (recessive), but two facing houses are merely opposite to one another (equi-relative). During these relationships, there may be motion of the things recited in the related items which we shall call "dynamic," i.e., the tractor-trailer example; or the things recited in the related items may be "static," i.e., both the book-table and the house-house examples.

By reserving the use of three binary columns, we can encode these relationships. In the first position we can put a 1 for static and a 0 for dynamic. The next two bits together can indicate either the dominant-recessive or the equirelative condition thusly:

dominant:	1	0
recessive:	0	1
equirelative:	0	0

SCHEDULE 14

This technique solves a specific problem in a chain sequence of conditions where, e.g., A drives B, B drives C, C drives D. With this type of disclosure, one might wish to retrieve C driven by A. Itemizing with this technique:

Item #		*Static-Motion Column*	*Dominant-Recessive Columns*	*Interfix*
503	A-(named-thing) Drive-(apparatus)	0	1 0	4
504	B-(named-thing) Drive-(apparatus)	0	1 1	4
505	C-(named-thing) Drive-(apparatus)	0	1 1	4
506	D-(named-thing) Drive-(apparatus)	0	0 1	4

SCHEDULE 15

we note that A is dominant only, B is recessive as to A but dominant as to C, etc. while D is recessive only as to C. By wording a question: "Find A with a dominant drive and C with a recessive drive," we can retrieve this portion of the disclosure, when the question is framed as stated above.

This technique is also adaptable in other situations. Consider the thing A which is taken from B to C and from C to D and from D to E. We can again use two adjacent columns with the code 1 0 as "from" for B and 0 1 as "to" for E, and the intermediate stations C and D would use the code 1 1 showing that they received the item "from" the item ahead and sent it "to" the item shown in Schedule 16.

This shorthand system cuts out the use of a concept column. However, if time was of the essence it could be handled with the interfixes assigned to the specific concepts, though each word involving both "from" and "to" would have to be repeated, as shown in Schedule 17.

Item #	*Descriptor and Modulant*	*From-To Columns*	*Interfix*
227	A-(work) Transport-(method)		4
228	B-(named-thing) Transport-(method)	1 0	4
229	C-(named-thing) Transport-(method)	1 1	4
230	D-(named-thing) Transport-(method)	1 1	4
231	E-(named-thing) Transport-(method)	0 1	4

SCHEDULE 16

Item #	*Descriptor and Modulant*	*Concept and Interfix*			*From-To Columns*
227	A-(work) Transport-(method)				
228	B-(named-thing) Transport-(method)	*Timafor-5*			1 0
229	C-(named-thing) Transport-(method)	*Timafor-5*			0 1
	Transport-(method)	*Timaft-5*	*Timafor-6*	4	1 0
230	D-(named-thing) Transport-(method)	*Timaft-5*	*Timafor-6*	4	0 1
	Transport-(method)		*Timaft-6*	4	1 0
231	E-(named-thing) Transport-(method)		*Timaft-6*	4	0 1

SCHEDULE 17

CONCLUSION

This constitutes the status of this project to date. Since the interrelational concepts now seem to be the most easy to derive, additional concepts and codes are now being worked on. Many other loose ends are clearly evident and need tying up. Much manpower and time are needed in deep research where, to date the surface has merely been scratched. Constructive criticism and comment from others will be most welcome.

REFERENCES

(1) Mechanized Searching in the U. S. Patent Office M. F. Bailey, B. E. Lanham, and J. Leibowitz, Journal of the Patent Office Society, Vol. 35, pp. 566-587.

Advances in Mechanization of Patent Searches B. E. Lanham, J. Leibowitz, and H. R. Koller Presented before Division of Chemical Literature 129th meeting of the American Chemical Society Dallas, Tex., April 11, 1956.

Comm--DC--43949

(2) William N. Locke and A. Donald Booth "Machine Translation of Languages," John Wiley & Sons--1955; pp. 167-173, inc.

(3) Classification Bulletin #402, U. S. Patent Office, 1951 containing the definitions of the classes and subclasses of class 53, Package Making.

James W. Perry, Allen Kent, and Madeline M. Berry "Machine Literature Searching" 1956, especially pages 84-89.

Since this paper went to press this new book has been received. The close similarity between our Modulants and the authors' Analytic and Synthetic Relationships is noted. The distinctions they draw between their two types of relationships do not, however, appear usable in the solution of the Patent Office problem. E.g., An "insect acted upon by an insecticide" (Analytic symbol W) is both a "starting material" (Synthetic code KAJ) and a "material processed" (Synthetic code KEJ) in Patent Office reasoning.